Bonn Summer School
Advances in Empirical Macroeconomics

Karel Mertens

Cornell, NBER, CEPR

Bonn, June 2015

In God we trust, all others bring data.

William E. Deming (1900-1993)

Angrist and Pischke are Mad About Macro

(2010 JEP, The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics)

Sims, JEP 2010, Comment on Angrist and Pischke: But economics is not an experimental science

Overview

1. Estimating the Effects of Shocks Without Much Theory

General Approach:

1. Using observables z_t, fit a model of expectations $E[z_t \mid \mathcal{I}_{t-1}]$
2. Identify meaningful shocks from innovations $z_t - E[z_t \mid \mathcal{I}_{t-1}]$.
3. Estimate dynamic causal effects of shocks/variance contributions.

$\mathcal{I}_t$: information available to economic decision makers at time t.

1.1 Structural Time Series Models

State Space (SS) Representation

The solution of stationary linear models can generally be written as

$$\begin{aligned} s_t &= \mathcal{G} s_{t-1} + \mathcal{F} e_t \\ z_t &= \mathcal{A} s_{t-1} + \mathcal{D} e_t \end{aligned}$$

s_t is a $m \times 1$ vector of state variables
e_{t+1} is an $l \times 1$ vector of uncorrelated white noise, or **structural shocks**

$$E[e_t] = 0 \ , \ E[e_t e_t'] = I \ , \ E[e_t e_s'] = 0 \text{ for } s \neq t$$

z_t is an $n \times 1$ vector of variables of interest
$\mathcal{G}$ is $m \times m$, $\mathcal{F}$ is $m \times l$, $\mathcal{A}$ is $n \times m$ and $\mathcal{D}$ is $n \times l$

Stability and Stationarity

The $m \times m$ matrix $\mathcal{G}$ has all eigenvalues less than one in modulus, i.e.

- $det(M - \lambda I) \neq 0$ for $| \lambda | \geq 1$, or equivalently
- $det(I - Mz) \neq 0$ for $| z | \leq 1$

s_t follows a stable VAR(1) process

s_t and z_t are **stationary stochastic processes**, i.e. the first and second moments are time invariant.

Lag Operator

Define the lag operator L, i.e. $L^k x_t = x_{t-k}$.

$$\begin{aligned} s_t &= \mathcal{G} s_{t-1} + \mathcal{F} e_t \\ (I - \mathcal{G} L) s_t &= \mathcal{F} e_t \\ s_t &= (I - \mathcal{G} L)^{-1} \mathcal{F} e_t \\ s_t &= \sum_{i=0}^{\infty} \mathcal{G}^i \mathcal{F} e_{t-i} \end{aligned}$$

Time Series Representations

Moving Average (MA) Representation

$$\text{MA(q)} \quad : \quad z_t = M(L)\upsilon_t = \sum_{i=0}^{q} \mathcal{M}_i \upsilon_{t-i}$$

where $M(L) = \mathcal{M}_0 + \mathcal{M}_1 L + ... + \mathcal{M}_q L^q$ and innovations process υ_t

$$E[\upsilon_t] = 0 \ , \ E[\upsilon_t \upsilon_t'] = \Sigma \ , \ E[\upsilon_t \upsilon_s'] = 0 \text{ for } s \neq t$$

Wold Representation Theorem
Every stationary process z_t can be written as an $MA(\infty)$.
(plus deterministic terms).

Stationary linear models for z_t and e_t can always be written as $MA(\infty)$

$$z_t \quad = \quad M^*(L)e_t = \sum_{i=0}^{\infty} \mathcal{M}_i^* e_{t-i}$$

Given an SS representation $\{\mathcal{G}, \mathcal{F}, \mathcal{A}, \mathcal{D}\}$,

$$\begin{aligned} z_t &= \sum_{i=1}^{\infty} \mathcal{A}\mathcal{G}^{i-1}\mathcal{F}e_{t-i} + \mathcal{D}e_t \\ &= \mathcal{A}(I - \mathcal{G}L)^{-1}\mathcal{F}e_{t-1} + \mathcal{D}e_t \\ &= \left(\mathcal{D} + \mathcal{A}(I - \mathcal{G}L)^{-1}\mathcal{F}L\right) e_t \end{aligned}$$

such that $\mathcal{M}_0^* = \mathcal{D}$ and $\mathcal{M}_i^* = \mathcal{A}\mathcal{G}^{i-1}\mathcal{F}$ for $i \geq 1$

Assume $n = l$ and **Stochastic Nonsingularity:**

$\mathcal{D}$ is an $n \times n$ invertible matrix.

We can write a 'structural' MA

$$z_t = M(L)\upsilon_t = \sum_{i=0}^{\infty} \mathcal{M}_i \upsilon_{t-i}$$

with $\upsilon_t = \mathcal{D}e_t$, $\Sigma = \mathcal{D}\mathcal{D}'$, $\mathcal{M}_0 = I$

$$z_t = \left(I + \mathcal{A}(I - \mathcal{G}L)^{-1}\mathcal{F}\mathcal{D}^{-1}L\right)\upsilon_t$$

such that $\mathcal{M}_i = \mathcal{A}\mathcal{G}^{i-1}\mathcal{F}\mathcal{D}^{-1}$ for $i \geq 1$

Time Series Representations

Vector Autoregressive Moving Average Representation

$$\text{VARMA(p,q)} \quad : \quad B(L)z_t = M(L)\upsilon_t$$

where

$$\begin{aligned} B(L) &= I - \mathcal{B}_1 L - \ldots - \mathcal{B}_p L^p \\ M(L) &= \mathcal{M}_0 + \mathcal{M}_1 L + \ldots + \mathcal{M}_q L^q \end{aligned}$$

$$E[\upsilon_t] = 0\ ,\ \ E[\upsilon_t \upsilon_t'] = \Sigma\ ,\ \ E[\upsilon_t \upsilon_s'] = 0 \text{ for } s \neq t$$

Starting from the MA representation of our linear models,

$$z_t = \mathcal{A}(I - \mathcal{G}L)^{-1}\mathcal{F}e_{t-1} + \mathcal{D}e_t$$

Suppose **n** $=$ **m** and $\mathcal{A}$ **is invertible**,

$$\begin{aligned} \mathcal{A}^{-1}z_t &= (I - \mathcal{G}L)^{-1}\mathcal{F}e_{t-1} + \mathcal{A}^{-1}\mathcal{D}e_t \\ (I - GL)\mathcal{A}^{-1}z_t &= \mathcal{F}e_{t-1} + (I - \mathcal{G}L)\mathcal{A}^{-1}\mathcal{D}e_t \end{aligned}$$

Assuming $\mathcal{D}$ invertible we obtain a structural VARMA(1,1),

$$z_t = \mathcal{A}\mathcal{G}\mathcal{A}^{-1}z_{t-1} + \upsilon_t - \mathcal{A}\left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right)\mathcal{A}^{-1}\upsilon_{t-1}$$

'Structural' here means

$$\upsilon_t = \mathcal{D}e_t$$

Time Series Representations

Vector Autoregressive Representation

$$\begin{aligned} \text{VAR(p)} \quad : \quad & B(L)z_t = \upsilon_t \\ & \Rightarrow z_t = \mathcal{B}_1 z_{t-1} + \ldots + \mathcal{B}_p z_{t-p} + \upsilon_t \end{aligned}$$

where

$$B(L) \quad = \quad I - \mathcal{B}_1 L - ... - \mathcal{B}_p L^p$$

$$E[\upsilon_t] = 0 \; , \; E[\upsilon_t \upsilon_t'] = \Sigma \; , \; E[\upsilon_t \upsilon_s'] = 0 \text{ for } s \neq t$$

Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007):

Start from the SS representation of our models,

$$\begin{aligned} s_t &= \mathcal{G}s_{t-1} + \mathcal{F}e_t \\ z_t &= \mathcal{A}s_{t-1} + \mathcal{D}e_t \end{aligned}$$

Assume $n = l$ and **Stochastic Nonsingularity:**

$\mathcal{D}$ is an $n \times n$ invertible matrix.

When $\mathcal{D}$ is nonsingular,

$$e_t = \mathcal{D}^{-1}(z_t - \mathcal{A}s_{t-1})$$

Substituting

$$s_t = \left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right) s_{t-1} + \mathcal{F}\mathcal{D}^{-1} z_t$$

Invertibility (in the Past):

The eigenvalues of $\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}$ are strictly less then one in modulus.

Under this condition we can write

$$s_t = \sum_{i=0}^{\infty} \left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right)^i \mathcal{F}\mathcal{D}^{-1} z_{t-i}$$

Substituting

$$z_t = \sum_{i=1}^{\infty} \mathcal{A}\left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right)^{i-1}\mathcal{F}\mathcal{D}^{-1} z_{t-i} + \mathcal{D}e_t$$

such that

$$\mathcal{B}_i = \mathcal{A}\left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right)^{i-1}\mathcal{F}\mathcal{D}^{-1}$$

$$\upsilon_t = \mathcal{D}e_t$$

In practice, lag truncation: $z_t = \sum_{i=1}^{p} \mathcal{B}_i z_{t-i} + \upsilon_t$

Stochastic Nonsingularity: No big deal (measurement errors)

Invertibility (in the Past): Choice of z_t is important!
If the invertibility condition does not hold and we estimate a VAR:

$$\upsilon_t \neq \mathcal{D}e_t$$

Alternatively, start from the MA representation

$$z_t = M(L)\upsilon_t$$

or VARMA representation

$$B(L)z_t = M(L)\upsilon_t$$

and invert $M(L)$ to obtain a VAR representation.

(Note: it is assumed that $M(L)$ is a square matrix of rational functions.)

$$S(L)B(L)z_t = S(L)M(L)v_t$$

$M(L)$ is **invertible in the past**,

i.e. there is an $S(L)$ such that $S(L)M(L) = I$ and $S(L)$ only has nonnegative powers of L,

if $\det(M(L)) \neq 0$ for $\mid L \mid \leq 1$.

See Hansen and Sargent (1980, 1991), Lippi and Reichlin (1993), Forni and Gambetti (2014)

MA representation of our models:

$$z_t = \left(I + \mathcal{A}(I - \mathcal{G}L)^{-1}\mathcal{F}\mathcal{D}^{-1}L\right) v_t$$

Using the matrix determinant lemma,

$$det\left(I + \mathcal{A}(I - \mathcal{G}L)^{-1}\mathcal{F}\mathcal{D}^{-1}L\right) = det\left(I - (\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A})L\right)/det(I - \mathcal{G}L)$$

The condition that

$$det\left(I - (\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A})L\right) \neq 0 \text{ for } \mid L \mid \leq 1$$

is equivalent to

$$det\left((\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}) - Iz\right) \neq 0 \text{ for } \mid z \mid \geq 1$$

or $\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}$ must have all eigenvalues strictly less then one in modulus (same as Fernandez-Villaverde et al. 2007).

VARMA representation of our models:

$$(I - \mathcal{A}\mathcal{G}\mathcal{A}^{-1}L)z_t \quad = \quad \left(I - \mathcal{A}\left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right)\mathcal{A}^{-1}L\right)\upsilon_{t-1}$$

The condition that

$$det\left(I - \mathcal{A}\left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right)\mathcal{A}^{-1}L\right) \neq 0 \text{ for } \mid L \mid \leq 1$$

is equivalent to

$$det\left(\left(\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}\right) - Iz\right) \neq 0 \text{ for } \mid z \mid \geq 1$$

or $\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A}$ must have all eigenvalues strictly less then one in modulus (same as Fernandez-Villaverde et al. 2007).

Example: New Keynesian Model

See Clarida, Gali and Gertler (1999) and the Woodford (2003) and Gali (2008) books

$$E_t \Delta \hat{y}^{gap}_{t+1} = \phi_\pi \pi_t - E_t \pi_{t+1} - u_t \qquad \text{(Eq. Euler)}$$
$$\pi_t = \kappa \hat{y}^{gap}_t + \beta E_t \pi_{t+1} - v_t \qquad \text{(Phillips curve)}$$

where $\kappa > 0$, $\phi_\pi > 1$, $0 \leq \beta < 1$

y^{gap}_t : output gap , π_t : inflation

v_t: cost push shocks

u_t: other shocks (technology, govt spending, taxes, monetary policy,...)

u_t and v_t are stationary exogenous processes

Iterating forward,

$$\begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} = E_t \sum_{j=0}^{\infty} \mathcal{C}^{-(j+1)} \begin{bmatrix} u_{t+j} \\ v_{t+j} \end{bmatrix}$$

where

$$\mathcal{C}^{-1} = \frac{1}{1+\phi_\pi \kappa} \begin{bmatrix} 1 & 1-\beta\phi_\pi \\ \kappa & \beta+\kappa \end{bmatrix}$$

Note $\mathcal{C}^{-1}$ has eigenvalues strictly less than one in modulus. ▸ Why?

Suppose the shocks follow a VAR(1) process.

$$\underbrace{\begin{bmatrix} u_t \\ v_t \end{bmatrix}}_{s_t} = \underbrace{\Lambda}_{\mathcal{G}} \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \underbrace{\Omega}_{\mathcal{F}} e_t$$

$$\begin{aligned} \underbrace{\begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix}}_{z_t} &= \sum_{j=0}^{\infty} \mathcal{C}^{-(j+1)} \Lambda^j \begin{bmatrix} u_t \\ v_t \end{bmatrix} \\ &= \underbrace{(\mathcal{C} - \Lambda)^{-1} \Lambda}_{\mathcal{A}} \underbrace{\begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix}}_{s_{t-1}} + \underbrace{(\mathcal{C} - \Lambda)^{-1} \Omega}_{\mathcal{D}} e_t \end{aligned}$$

Note that $\mathcal{G} - \mathcal{F}\mathcal{D}^{-1}\mathcal{A} = 0$.

State Space representation:

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \Lambda \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \Omega e_t$$

$$\begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} = (\mathcal{C} - \Lambda)^{-1} \Lambda \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + (\mathcal{C} - \Lambda)^{-1} \Omega e_t$$

Moving Average Representation

$$\begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} = \sum_{i=0}^{\infty} (\mathcal{C} - \Lambda)^{-1} \Lambda^i \Omega e_{t-i}$$

VAR/VARMA Representation

$$\begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} = (\mathcal{C} - \Lambda)^{-1} \Lambda (\mathcal{C} - \Lambda) \begin{bmatrix} \hat{y}_{t-1}^{gap} \\ \pi_{t-1} \end{bmatrix} + (\mathcal{C} - \Lambda)^{-1} \Omega e_t$$

Reduced Form Parameters

$$
\begin{aligned}
SS : s_t &= \mathcal{G} s_{t-1} + \mathcal{H} \upsilon_t && \mathcal{G}, \mathcal{A}, \mathcal{H}, \Sigma \\
z_t &= \mathcal{A} s_{t-1} + \upsilon_t && \\
\text{MA(q)} \;:\; z_t &= M(L) \upsilon_t && \mathcal{M}_i, \Sigma \\
\text{VARMA(p,q)} :\; B(L) z_t &= M(L) \upsilon_t && \mathcal{B}_i, \mathcal{M}_i, \Sigma \\
\text{VAR(p)} :\; B(L) z_t &= \upsilon_t && \mathcal{B}_i, \Sigma
\end{aligned}
$$

where $E[\upsilon_t] = 0$, $E[\upsilon_t \upsilon_t'] = \Sigma$, $E[\upsilon_t \upsilon_s'] = 0$ for $s \neq t$

Note: SS, MA and VARMA require additional normalizations.

When well-specified, these are all models that allow us to separate expectations $E[z_t \mid \mathcal{I}_{t-1}]$ and innovations $\upsilon_t = z_t - E[z_t \mid \mathcal{I}_{t-1}]$.

Estimation of Reduced Form Parameters

▸ VAR estimation

Some references:

- Hamilton, 1994, 'Time Series Analysis'
- Luetkepohl, 2005, 'A New Introduction to Time Series Analysis'
- Brockwell and Davis, 2006,'Time Series: Theory and Methods'
- Aoki, 1990, 'State Space Modeling of Time Series'
- Durbin and Koopman, 2012, 'Time Series Analysis by State Space Methods'

Local Uniquess

In matrix form

$$E_t \begin{bmatrix} \hat{y}_{t+1}^{gap} \\ \pi_{t+1} \end{bmatrix} = \mathcal{C} \begin{bmatrix} \hat{y}_t^{gap} \\ \pi_t \end{bmatrix} - \begin{bmatrix} u_t \\ v_t \end{bmatrix} \quad , \quad \mathcal{C} \equiv \frac{1}{\beta} \begin{bmatrix} \beta + \kappa & \beta\phi_\pi - 1 \\ -\kappa & 1 \end{bmatrix}$$

The companion matrix $\mathcal{C}$ has the characteristic polynomial

$$\begin{aligned} \mathcal{P}(\varphi) &= \varphi^2 - \text{tr}(\mathcal{C})\varphi + \det(\mathcal{C}) \\ \text{tr}(\mathcal{C}) &= 1 + 1/\beta + \kappa/\beta > 1 \\ \det(\mathcal{C}) &= (1 + \kappa\phi_\pi)/\beta > 1 \end{aligned}$$

which has roots outside the unit circle if $\text{tr}(\mathcal{C}) < 1 + \det(\mathcal{C})$ or

$$\phi_\pi > 1 \qquad \text{(Taylor Principle)}$$

Back

Estimating a VAR

Sample of $T + p$ observations of an $n \times 1$ vector z_t:

$$\{z_{t-p+1}, z_{t-p+2}, \ldots, z_{T-1}, z_T\}$$

Define the $n \times T$ matrix z such that:

$$z = \begin{bmatrix} z_1 & z_2 & \cdots & z_T \end{bmatrix}$$

Define a $np \times 1$ vector Z_t:

$$Z_t = \begin{bmatrix} z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-p} \end{bmatrix}$$

Let Z be a $np \times T$ matrix collecting T observations of Z_t:

$$Z = \begin{bmatrix} Z_1 & Z_2 & \cdots & Z_p \end{bmatrix}$$

Let υ be a $n \times T$ matrix of $n \times 1$ residuals υ_t:

$$\upsilon = \begin{bmatrix} \upsilon_1 & \upsilon_2 & \cdots & \upsilon_T \end{bmatrix}$$

Let B be a $n \times np$ matrix of coefficients:

$$B = \begin{bmatrix} B_1 & B_2 & \cdots & B_p \end{bmatrix}$$

Introduce the vectorization operator:

$$\mathbf{z} = \mathrm{vec}(z)$$
$$\mathbf{u} = \mathrm{vec}(\upsilon)$$

Where $\mathbf{z}$ is a $nT \times 1$ vector of the stacked columns of z. The variance-covariance matrix of $\mathbf{u}$ is:

$$\mathrm{Var}(\upsilon) \equiv \mathbf{\Sigma} = I_T \otimes \Sigma$$

Generalized Least Squares

Re-write the reduced form VAR(p) as:

$$z = BZ + \mathbf{u}$$

Or as:

$$\mathbf{z} = (Z' \otimes I_n)\beta + \mathbf{u}$$

Where $\otimes$ is the Kronecker product. We can estimate β with Generalized Least Squares (GLS):

$$\begin{aligned}
\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u} &= (\mathbf{z} - (Z' \otimes I_n)\beta)'\mathbf{\Sigma}^{-1}(\mathbf{z} - (Z' \otimes I_n)\beta) \\
&= \mathbf{z}'\mathbf{\Sigma}^{-1}\mathbf{z} + \beta'(Z \otimes I_n)\mathbf{\Sigma}^{-1}(Z' \otimes I_n)\beta - 2\beta'(Z \otimes I_n)\mathbf{\Sigma}^{-1}\mathbf{z} \\
&= \mathbf{z}'(I_T \otimes \Sigma^{-1})\mathbf{z} + \beta'(ZZ' \otimes \Sigma^{-1})\beta - 2\beta'(Z \otimes \Sigma^{-1})\mathbf{z}
\end{aligned}$$

First order condition:

$$2(ZZ' \otimes \Sigma^{-1})\beta - 2(Z \otimes \Sigma^{-1})\mathbf{z} = 0$$

The GLS estimator is therefore:

$$\hat{\beta} = ((ZZ')^{-1}Z \otimes I_n)\mathbf{z}$$

This is the same as OLS or ML.

Asymptotic normality:

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma)$$

where $\Gamma = \mathrm{plim} ZZ'/T$ and the estimators are

$$\hat{\Gamma} = \frac{ZZ'}{T}$$

$$\hat{\Sigma} = \frac{1}{T - np - 1} z(I_T - Z'(ZZ')^{-1}Z)z'$$